# Xinya Du

| CONTACT | Assistant Professor | Email: xinya.du@utdallas.edu |
|---|---|---|
| | Department of Computer Science | Website: https://xinyadu.github.io |
| | The University of Texas at Dallas | Google Scholar |

**RESEARCH INTERESTS**

Natural Language Processing, Machine Learning, Large Language Models, AI for Science.

**EDUCATION**

**Cornell University** — Aug 2016 – Aug 2021
Ph.D. in Computer Science (M.S. degree granted in Aug 2019)
Advisor: Claire Cardie

**Shanghai Jiao Tong University** — Sep 2012 – Aug 2016
B.E. in Computer Science and Engineering
Outstanding Graduates Award

**PROFESSIONAL EXPERIENCE**

**University of Texas at Dallas** — Aug 2022 – Present
*Assistant Professor in Computer Science*

**University of Illinois at Urbana-Champaign** — Sep 2021 – Aug 2022
*Postdoctoral Research Associate. Advisor: Heng Ji*

**INDUSTRIAL EXPERIENCE**

**Google AI Research** — May 2020 – Aug 2020
*Research Intern*

**Allen Institute for Artificial Intelligence (AI2)** — Sep 2018 – Dec 2018
*Research Intern*

**Microsoft Research (MSR)** — May 2018 – Aug 2018
*Research Intern*

**SELECTED AWARDS & HONORS**

| | |
|---|---|
| **NSF CAREER award** | National Science Foundation, 2024 |
| **AAAI New Faculty Highlights** | AAAI, 2024 |
| **Cisco Faculty Research Award** | Cisco, 2024 |
| **Best Poster Award** (Top 1%) | AI for Science Workshop, 2024 |
| **Amazon Research Award** | Amazon, 2023 |
| **Spotlight Rising Star in Data Science** | University of Chicago, 2021 |
| **Top 100 New Stars in Artificial Intelligence** | Baidu Scholar, 2020 |
| **Most Influential ACL Papers** (15 each year) | Paperdigest, 2017 |
| **National Scholarship** (Top 1% students nationwide) | SJTU, 2013 |

GRANTS   **NSF CAREER: Learning to Extract Consistent Event Graphs from Long and Complex Documents (PI)**
Funding Source: National Science Foundation.
Period: May 2024 – present.
Amount Awarded: $561,219. Amount to me: $561,219.
**Awarded on first submission.**

**Process-guided Fine-tuning for Answering Complex Questions (PI)**
Funding Source: Amazon Research Award.
Period: Jan 2024 – present.
Amount Awarded: $110,000 ($40,000 AWS credits). Amount to me: $110,000.

**Amazon Trusted AI Challenge Grant (Co-PI)**
Funding Source: Amazon.
Period: Jan 2025 – present.
Amount Awarded to Our Team: $250,000.

**FAIGen: Faithful LLM Generation with Scientific Principles-guided Learning (PI)**
Funding Source: Cisco Faculty Research Award.
Period: Jan 2025 – present.
Amount Awarded: $50,000. Amount to me: $50,000.

**Faithfulness in Large Vision Language Models (PI)**
Funding Source: OpenAI Researcher Access Program.
Period: Sep 2024 – present.
Amount Awarded: $4,000. Amount to me: $4,000.

**Undergraduate Research Apprenticeship Award (URAP) (PI)**
Funding Source: University of Texas at Dallas.
Period: Summer 2023, 2024.
Amount Awarded: $9,000. Amount to me: $9,000.

PUBLICATIONS   Note: $^\dagger$ indicates that I am a co-leading author. $^*$ indicates equal contributions.
You can also find my publication list on [DBLP] and [Google Scholar] pages.

[1]   **PRD: Peer Rank and Discussion Improve Large Language Model-based Evaluations** [pdf]
Ruosen Li, Teerth Patel, **Xinya Du**
*In Transactions on Machine Learning Research (TMLR), 2024.*

[2]   **Large Language Models for Automated Open-domain Scientific Hypotheses Discovery** [pdf]
Zonglin Yang, **Xinya Du**$^\dagger$, Junxian Li, Jie Zheng, Soujanya Poria, Erik Cambria
*In Findings of the Association for Computational Linguistics: (ACL), 2024.*
**Best poster award in AI4Science workshop, 2024 (Top 1%).**

[3] **FaithScore: Fine-grained Evaluations of Hallucinations in Large Vision-Language Models** [pdf]
Liqiang Jing, Ruosen Li, Yunmo Chen, **Xinya Du**
*In Findings of the Association for Computational Linguistics: (EMNLP), 2024.*

[4] **IQA-EVAL: Automatic Evaluation of Human-Model Interactive Question Answering** [pdf]
Ruosen Li, Ruochen Li, Barry Wang, **Xinya Du**
*In Annual Conference on Neural Information Processing Systems (NeurIPS), 2024.*

[5] **QAEvent: Event Extraction as Question-Answer Pairs Generation** [pdf]
Milind Choudhary, **Xinya Du**
*In Findings of the Association for Computational Linguistics: (EACL), 2024.*

[6] **Language Models as Inductive Reasoners** [pdf]
Zonglin Yang, Li Dong, **Xinya Du**[†], Hao Cheng, Erik Cambria, Xiaodong Liu, Jianfeng Gao, Furu Wei
*In Conference of the European Chapter of the Association for Computational Linguistics (EACL), 2024.*

[7] **Making Natural Language Reasoning Explainable and Faithful** [pdf]
**Xinya Du**
*In AAAI Conference on Artificial Intelligence (AAAI), 2024.*

[8] **Proto-CLIP: Vision-Language Prototypical Network for Few-Shot Learning** [pdf]
Jishnu Jaykumar P, Kamalesh Palanisamy, Yu-Wei Chao, **Xinya Du**, Yu Xiang
*In International Conference on Intelligent Robots and Systems (IROS), 2024.*

[9] **Document-level Causal Relation Extraction with Knowledge-guided Binary Question Answering** [pdf]
Zimu Wang, Lei Xia, Wei Wang, **Xinya Du**
*In Findings of the Association for Computational Linguistics: (EMNLP), 2024.*

[10] **Leveraging Structured Information for Explainable Multi-hop Question Answering and Reasoning** [pdf]
Ruosen Li, **Xinya Du**
*In Findings of the Association for Computational Linguistics: (EMNLP), 2023.*

[11] **Process of Elimination for Multiple Choice Reasoning** [pdf]
Chenkai Ma, **Xinya Du**
*In Conference on Empirical Methods in Natural Language Processing (EMNLP), 2023.*

[12] **Probing Representations for Document-level Event Extraction** [pdf]
Barry Wang, **Xinya Du**, Claire Cardie
*In Findings of the Association for Computational Linguistics: (EMNLP), 2023.*

[13] **End-to-end Case-Based Reasoning for Commonsense Knowledge Base Completion** [pdf]
Zonglin Yang, **Xinya Du**[†], Erik Cambria, Claire Cardie
*In Conference of the European Chapter of the Association for Computational Linguistics*

*(EACL), 2023.*

[14] **Zero-Shot Classification by Logical Reasoning on Natural Language Explanations** [pdf]
Chi Han, Hengzhi Pei, **Xinya Du**, Heng Ji
*In Findings of the Association for Computational Linguistics: (ACL), 2023.*

[15] **Toward Consistent and Informative Event-Event Temporal Relation Extraction** [pdf]
Xiaomeng Jin, Haoyang Wen, **Xinya Du**, Heng Ji *In MATCHING at Annual Meeting of the Association for Computational Linguistics (ACL), 2023.*

[16] **Logical Entity Representation in Knowledge-Graphs for Differentiable Rule Learning** [pdf]
Chi Han, Qizheng He, Charles Yu, **Xinya Du**, Hanghang Tong, Heng Ji
*In International Conference on Learning Representations (ICLR), 2023.*

[17] **RESIN-11: Schema-guided Event Prediction for 11 Newsworthy Scenarios** [pdf]
**Xinya Du**, Zixuan Zhang, Sha Li, Heng Ji, the RESIN team
*In Conference of the North American Chapter of the Association for Computational Linguistics (NAACL): System Demonstrations, 2022.*
**Top ranking system in DARPA KAIROS evaluation. [Link]**

[18] **Retrieval-Augmented Generative Question Answering for Event Argument Extraction** [pdf]
**Xinya Du**, Heng Ji
*In Conference on Empirical Methods in Natural Language Processing (EMNLP), 2022.*

[19] **Dynamic Global Memory for Document-level Argument Extraction** [pdf]
**Xinya Du**, Sha Li, Heng Ji
*In Annual Meeting of the Association for Computational Linguistics (ACL), 2022.*

[20] **Automatic Error Analysis for Document-level Information Extraction** [pdf]
Aliva Das*, **Xinya Du***, Barry Wang*, Kejian Shi, Jiayuan Gu, Thomas Porter, Claire Cardie
*In Annual Meeting of the Association for Computational Linguistics (ACL), 2022.*

[21] **Template Filling with Generative Transformers** [pdf]
**Xinya Du**, Alexander M. Rush, Claire Cardie
*In Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2021.*

[22] **GRIT: Generative Role-filler Transformers for Document-level Event Entity Extraction** [pdf]
**Xinya Du**, Alexander M. Rush, Claire Cardie
*In Conference of the European Chapter of the Association for Computational Linguistics (EACL), 2021.*

[23] **Few-shot Intent Classification and Slot Filling with Retrieved Examples** [pdf]
Dian Yu, Luheng He, Yuan Zhang, **Xinya Du**, Panupong Pasupat, Qi Li
*In Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2021.*

[24] **QA-Driven Zero-shot Slot Filling with Weak Supervision Pretraining** [pdf]
**Xinya Du**, Luheng He, Qi Li, Dian Yu, Panupong Pasupat, Yuan Zhang
*In Annual Meeting of the Association for Computational Linguistics (ACL), 2021.*

[25] **Event Extraction by Answering (Almost) Natural Questions** [pdf]
**Xinya Du**, Claire Cardie
*In Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020.*
**Top 1% most cited articles published in Computer Science in 2020.**

[26] **Improving Event Duration Prediction via Time-aware Pre-training** [pdf]
Zonglin Yang, **Xinya Du**, Alexander M. Rush, Claire Cardie
*In Findings of the Association for Computational Linguistics: (EMNLP), 2020.*

[27] **Document-Level Event Role Filler Extraction using Multi-Granularity Contextualized Encoding** [pdf]
**Xinya Du**, Claire Cardie
*In Annual Meeting of the Association for Computational Linguistics (ACL), 2020.*

[28] **Leveraging Structured Metadata for Improving Question Answering on the Web** [pdf]
**Xinya Du**, Adam Fourney, Robert Sim, Claire Cardie, Paul Bennett, Ahmed Hassan Awadallah
*In Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (AACL/IJCNLP), 2020.*

[29] **Be Consistent! Improving Procedural Text Comprehension using Label Consistency** [pdf]
**Xinya Du**, Bhavana Dalvi, Niket Tandon, Antoine Bosselut, Wen-tau Yih, Peter Clark, Claire Cardie
*In Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2019.*

[30] **Harvesting Paragraph-Level Question-Answer Pairs from Wikipedia** [pdf]
**Xinya Du**, Claire Cardie
*In Annual Meeting of the Association for Computational Linguistics (ACL), 2018.*
**Top 1% most cited articles published in Computer Science in 2018.**

[31] **Identifying Where to Focus in Reading Comprehension for Neural Question Generation** [pdf]
**Xinya Du**, Claire Cardie
*In Conference on Empirical Methods in Natural Language Processing (EMNLP), 2017.*

[32] **Learning to Ask: Neural Question Generation for Reading Comprehension** [pdf]
**Xinya Du**, Junru Shao, Claire Cardie
*In Annual Meeting of the Association for Computational Linguistics (ACL), 2017.*
**Most Influential ACL Papers (15 each year). [Link]**
**Featured in New Scientist and TechRepublic. [Link]**
**Top 0.1% most cited articles published in Computer Science in 2017.**

[33] **Cornell Belief and Sentiment System at TAC 2016** [pdf]
Vlad Niculae, Kai Sun, Xilun Chen, Yao Cheng, **Xinya Du**, Esin Durmus, Arzoo Katiyar,
Claire Cardie
*In Text Analysis Conference (TAC), 2016.*

MENTORING
EXPERIENCE

**PhD students**

Ruosen Li (2022–present, UT Dallas PhD student).
Publications: TMLR 2024, NeurIPS 2024, EMNLP Findings 2023 2024.

Liqiang Jing (2023–present, UT Dallas PhD student).
Publications: EMNLP Findings 2024.
Award: OpenAI Researcher Access Program.

Ruochen Li (2023–present, UT Dallas PhD student).
Publications: NeurIPS 2024.

Guiming Chen (2024–present, UT Dallas PhD student).

**Master Students**

Milind Choudhary (2023, UT Dallas Master student) → UT Dallas PhD.
Topic: Event Extraction as Question-Answer Pairs Generation.
Publications: EACL 2024 Findings.

Zimu Wang (2023–present, UT Dallas Visiting student) → UoL PhD.
Topic: Temporal relation extraction.
Publications: NeurIPS 2024, EMNLP 2024.

Son Tran (2023, UT Dallas Visiting student) → Cornell PhD.
Topic: Question Answering.
Publications: NeurIPS 2024.

Chenkai Ma (2023, UT Dallas Visiting student).
Topic: Multiple Choice Questions Reasoning.
Publications: EMNLP 2023.

Barry Wang (2021–2024, Cornell CS undergraduate student) → CMU PhD.
Topic: Automatic Error Analysis for Information Extraction.
Publications: ACL 2022, SciNLP 2022, EMNLP 2023, NeurIPS 2024.

Zonglin Yang (2020–2024, Cornell CS MEng student) → NTU PhD.
Topic: Commonsense and Case-based Reasoning for NLP.
Publications: EMNLP 2020 Findings, EACL 2023, ACL 2024.

Ehsan Aghazadeh (2024, UT Dallas Visiting student).
Topic: Large vision language models interpretability.

Bowen Yan (2024, UT Dallas Visiting student).
Topic: Large vision language models hallucinations.

Zhengsong Zhang (2024, UT Dallas Visiting student).
Topic: Large vision language models hallucinations.

Ziming Luo (2024, UT Dallas Visiting student).
Topic: Large language model for scientific discovery.

**Undergraduate/High School Students**

Teerth Patel (UT Dallas BS student, 2023–present).
Topic: Large language model peer evaluations.
Publications: TMLR 2024.
Award: Jonsson School of Engineering and Computer Science Award.

Lei Xia (2023–2024, UT Dallas Visiting student).
Topic: Multi-hop Question Answering.
Publications: NeurIPS 2024.

Minhao Zou (2024, UT Dallas Visiting student).
Topic: LLM for Scientific Discovery.

Arjun Junghare (2024, UT Dallas CS undergraduate student).
Topic: Image Generation with Large Language Models.

Shreya Kumara (2024, President of UTD Society of Asian Scientists and Engineers).

Anjana Bharadwaj (2024, President of UTD Women Who Compute).

Jaden Nunes (Summer 2023, DFW Local K-12 student).
Topic: Event Extraction as Question-Answer Pairs Generation.

Rishab Bhattacharya (Summer 2023, DFW Local K-12 student).
Topic: Event Extraction as Question-Answer Pairs Generation.

Shreyas Kumar (Summer 2023, DFW Local K-12 student).
Topic: Event Extraction as Question-Answer Pairs Generation.

Aliva Das (2021–2022, Cornell CS undergraduate student) → Amazon.
Topic: Automatic Error Analysis for Information Extraction.
Publications: ACL 2022, SciNLP 2021.

Rishi Malhotra (Spring 2021, Cornell CS undergraduate student) → Microsoft.
Topic: Applying Neural Document-level IE Model to the Scientific Domain.

Maitreyi Chatterjee (Spring 2021, Cornell CS undergraduate student) → LinkedIn.
Topic: Applying Neural Document-level IE Model to the Scientific Domain.

TEACHING
EXPERIENCE

Introduction to Machine Learning, UT Dallas, Spring 2024.

Natural Language Processing, UT Dallas, Fall 2023, 2024.

Deep Learning for Natural Language Processing, UT Dallas, Spring 2023.
**New course developed and taught (overall Instructor Score of 4.75/5.0). [Link]**

Natural Language Processing, UT Dallas, Fall 2022.

Natural Language Processing, Cornell University, Fall 2019.
Teaching Assistant for Prof. Claire Cardie.

Natural Language Processing, Cornell University, Spring 2019.
    Teaching Assistant for Prof. Yoav Artzi.

Software Engineering, Cornell University, Spring 2017, 2018.
    Teaching Assistant for Prof. William Arms.

Introduction to Computing Using Python, Cornell University, Fall 2016.
    Teaching Assistant for Prof. Walker White.

PROFESSIONAL **Chairing Roles:**
SERVICES
ACL (Annual Meeting of the Association for Computational Linguistics) Rolling Review Area Chair, 2024.

ACL Area Chair, 2023.

ACL Session Chair, 2024.

EMNLP (Conference on Empirical Methods in Natural Language Processing) Area Chair, 2024.

EMNLP Demo Track Area Chair, 2024.

NAACL (Conference of the North American Chapter of the Association for Computational Linguistics) Website Chair, 2024.

COLING (International Conference on Computational Linguistics) Senior Area Chair, 2024, 2025.

**Seminar/Conference/Workshop Organizing:**

AI4Research: Towards a Unified Knowledge-grounded Scientific Research Lifecycle.
Organizing team: Qingyun Wang, Wenpeng Yin, Lifu Huang, Yi R. Fung, Xinya Du, Carl Edwards, Tom Hope.
To Appear in AAAI 2025.

**Journal Reviewer:**

TKDE (IEEE Transactions on Knowledge and Data Engineering).

TASLP (IEEE Transactions on Audio, Speech and Language Processing).

TNNLS (IEEE Transactions on Neural Networks and Learning Systems).

TALLIP (ACM Transactions on Asian and Low-Resource Language Information Processing).

TKDD (ACM Transactions on Knowledge Discovery from Data).

CL (Computational Linguistics).

KAIS (Knowledge and Information Systems).

AI Communication.

IPM (Information Processing and Management).

**Conference Committee Member:**

ACL (Annual Meeting of the Association for Computational Linguistics)

NeurIPS (Conference on Neural Information Processing Systems)

ICLR (International Conference on Learning Representations)

AISTATS (International Conference on Artificial Intelligence and Statistics)

EMNLP (Conference on Empirical Methods in Natural Language Processing)

NAACL (Conference of the North American Chapter of the Association for Computational Linguistics)

AACL/IJCNLP (Asia-Pacific Chapter of the Association for Computational Linguistics / International Joint Conference on Natural Language Processing)

IJCAI (International Joint Conference on Artificial Intelligence)

AAAI (AAAI Conference on Artificial Intelligence)

NLPCC (Natural Language Processing and Chinese Computing)

CoNLL (Conference on Computational Natural Language Learning)

WNUT (Workshop on Noisy User-generated Text)

MRQA (Workshop on Machine Reading for Question Answering)

SEM (Joint Conference on Lexical and Computational Semantics)

**PhD/MS Committee Member:**

PhD: Xiangci Li, Jishnu Jaykumar Padalunkal, Basel Abdeen, Yibo Hu.

MS: Wooseong Yang, Shubham Patel

**Other Activities:**

Faculty advisor and judge: Association for Computing Machinery (ACM) Symposium, UT Dallas, 2023–present.

Faculty Advisor: Women Who Compute (WWC), UT Dallas, 2024–present.

Faculty Advisor: Society of Asian Scientists and Engineers (SASE), UT Dallas, 2024 – present.

Faculty Host: K-12 Outreach, UT Dallas, Summer 2023, 2024.

Research Mentor: Research, Inquiry, Design Experience (RIDE) Project, UT Dallas, 2024.

PhD Admission Committee Member, UT Dallas CS Department, 2022, 2023, 2024.

Site Host: North American Computational Linguistics Olympiad (NACLO), 2023, 2024.

PhD Admission Committee Member, Cornell CS Department, 2021.

Volunteer: Cornell CS Department PhD Visit Day, 2019, 2020, 2021.

Student Volunteer: ACL, 2017, 2018.

Student Volunteer: EMNLP, 2017.

RECENT TALKS

Synergizing Knowledge and Large Language Models.
University of Illinois Urbana-Champaign, Data Mining Group Seminar, Oct 2024.
University of Massachusetts–Amherst, Machine Learning & Friends Lunch, Sep 2024.

Synergy between Large Language Model and Knowledge.
University of North Texas, Oct 2024.

Faculty Round Table Talk.
UT Dallas Hobson Wildenthal Honors College, Aug 2024.

Large Language Models: Knowledge, Reasoning and Factuality.
Samsung Electronics America, Mar 2024.

Open-ended Evaluations of Foundational Models: Alignment and Faithfulness.
  Shanghai Jiao Tong University, Computer Science Department Seminar, Dec 2023.
  Fudan University, Computer Science Department Seminar, Dec 2023.

ChatGPT: Fact vs. Fiction.
  UT Dallas, Forum sponsored by The Dallas Morning News, May 2023.

Towards More Intelligent Extraction of Information from Documents.
  Allen Institute for Artificial Intelligence, Feb 2022.
  Simon Fraser University, Feb 2022.
  Rensselaer Polytechnic Institute, Feb 2022.
  University of California, Merced, Feb 2022.
  Hong Kong University of Science and Technology, February 2022.

Towards More Intelligent Extraction of Information from Documents.
  UIUC, Siebel School of Computing and Data Science Speaker Series, Feb 2022.

Towards More Informed Extraction of Events from Documents.
  University of Chicago, Rising Stars in Data Science Workshop, Jan 2021.
  Tencent AI Research America, Nov 2020.

Event Extraction by Answering (Almost) Natural Question.
  UIUC, Information Extraction and Knowledge Acquisition Class, Sep 2020.

LwLL: Progress on the NLP Front.
  Cornell University, DARPA site visit, Apr 2020.

Harvesting Paragraph-Level Question-Answer Pairs from Wikipedia.
  56th Annual Meeting of the Association for Computational Linguistics, July 2018.